Metabarcoding and Metagenomics 10: 357–375 (2026)
DOI: 10.3897/mbmg.10.162895

Methods

Collection of ITS2 region reference sequences for freshwater fish in the Japanese archipelago for environmental DNA metabarcoding applications

Gen Ito[1], Daisuke Sasaki[2], Masayuki K. Sakata[2,3], Seigo Kawase[4], Toshifumi Minamoto[2], Hiroki Yamanaka[1]

1 *Center for Biodiversity Science, Ryukoku University, 1-5 Yokotani, Seta Oe-cho, Otsu, Shiga 520-2194, Japan*
2 *Graduate School of Human Development and Environment, Kobe University, 3-11 Tsurukabuto, Nada-Ku, Kobe, Hyogo 657-8501, Japan*
3 *Research Faculty of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-Ku, Sapporo, Hokkaido 060-8589, Japan*
4 *Lake Biwa Museum, 1091 Oroshimo, Kusatsu, Shiga 525-0001, Japan*
Corresponding author: Gen Ito (sakurahayabusa6647@gmail.com)

Abstract

To enhance taxonomic resolution in environmental DNA (eDNA) metabarcoding, we evaluated the utility of the internal transcribed spacer 2 (ITS2) region. We developed ITS2 metabarcoding primers and constructed an ITS2 reference database for Japanese freshwater fishes by sequencing an additional 61 species and compiling existing sequences. We then compared the taxonomic resolution of ITS2 metabarcoding with that of the widely used mitochondrial 12S rRNA (MiFish) marker. Our results demonstrated that ITS2 provided higher resolution, particularly for species in *Oncorhynchus* and *Rhinogobius*, which are often difficult to distinguish using MiFish. The improved resolution in *Oncorhynchus* highlights ITS2's potential for monitoring economically and ecologically important salmonid species. Additionally, ITS2 resolved some closely related species that MiFish could not differentiate, underscoring its value in biodiversity assessments. Our findings suggest that ITS2 can complement MiFish in eDNA-based fish biodiversity monitoring, particularly for taxa where mitochondrial markers have limited resolution. Future studies should further explore the applicability of ITS2 across diverse freshwater ecosystems to optimize its integration into multi-marker metabarcoding approaches.

Key words: Biodiversity, Lake Biwa, MiFish, nuclear DNA, *Oncorhynchus*

OPEN ACCESS

Academic editor: Ana Filipa Filipe
Received: 23 June 2025
Accepted: 2 March 2026
Published: 2 April 2026

Citation: Ito G, Sasaki D, Sakata MK, Kawase S, Minamoto T, Yamanaka H (2026) Collection of ITS2 region reference sequences for freshwater fish in the Japanese archipelago for environmental DNA metabarcoding applications. Metabarcoding and Metagenomics 10: e162895. https://doi.org/10.3897/mbmg.10.162895

Copyright: © Gen Ito et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

The degradation of freshwater ecosystems due to anthropogenic impacts and climate change is a serious concern, making the conservation of freshwater fish, which are key representatives of these ecosystems, an urgent priority (Barbarossa et al. 2021; Su et al. 2021). To conserve freshwater fish, it is essential to rapidly obtain information on their spatial distributions (Kano et al. 2016). However, this requires a variety of specialized sampling tools and techniques tailored to the fish habitats (Huang et al. 2013), making it difficult to obtain accurate distribution data.

Environmental DNA (eDNA), which enables the detection of organisms' DNA from water samples without direct capture, has emerged as a powerful tool with greater sensitivity than conventional survey methods for assessing species distributions (Sigsgaard et al. 2015; Hänfling et al. 2016; Yamamoto et al. 2017). In particular, eDNA metabarcoding enables the simultaneous detection of DNA from multiple species (e.g., fish, birds, mammals) in a single water sample (Miya et al. 2015; Ushio et al. 2017, 2018). For these reasons, eDNA metabarcoding is a powerful method for rapidly assessing species' spatial distributions.

Currently, mitochondrial DNA (mtDNA) markers are the predominant choice for eDNA metabarcoding of fish. The high copy number of mtDNA per cell and its extensive use in molecular phylogenetics make it an effective target for eDNA studies (Alberts et al. 2002; Iwasaki et al. 2013). One of the most widely used primer sets, MiFish (Miya et al. 2015), which targets the mitochondrial 12S rRNA gene, has been applied across diverse aquatic environments, including marine, riverine, and lacustrine systems (Hänfling et al. 2016; Yamamoto et al. 2017; Ito et al. 2023). However, despite the evolutionary rate of mtDNA, species-level discrimination can sometimes be difficult, particularly among closely related taxa. In addition, its maternal inheritance prevents the detection of hybrids (Avise et al. 1979).

Although nuclear DNA (nDNA) is generally considered less suitable for eDNA analysis due to its low copy number per cell, ribosomal DNA (rDNA) exist in multiple copies and exhibit high interspecific variation, making them promising candidates for distinguishing closely related species (Long and Dawid 1980; Booton et al. 1999). Internal transcribed spacer (ITS) regions, which are part of rDNA, have been shown to exhibit higher eDNA detection sensitivity compared to mtDNA in studies on common carp (*Cyprinus carpio*) and Japanese giant salamander (*Andrias japonicus*) (Minamoto et al. 2017; Hidaka et al. 2024). Although these findings are based on species-specific analyses, they suggest that ITS regions could serve as superior genetic markers for eDNA metabarcoding and may complement or even surpass mtDNA-based approaches. Developing an eDNA metabarcoding framework utilizing nuclear markers may enhance species detection rates and improve taxonomic resolution. Additionally, estimating reproductive activity based on the nuclear-to-mitochondrial DNA ratio, which has traditionally been applied on a species-specific basis, could be extended to assess multiple species simultaneously (Wu et al. 2022).

Despite these advantages, the application of ITS regions in fish molecular phylogenetics remains limited (Booton et al. 1999), and no regionally comprehensive reference database has been established. To improve species identification accuracy in eDNA metabarcoding, it is essential to design ITS metabarcoding primers specifically for fish, sequence ITS regions for all species within a target region and construct a localized reference library.

In this study, we aimed to develop an ITS-based eDNA metabarcoding approach by designing primers and collecting reference sequences. As an initial step, we focused on the Lake Biwa basin, which hosts the highest freshwater fish diversity in the Japanese archipelago. The fish fauna of Lake Biwa is well-documented (Fujioka et al. 2024), and an established MiFish reference database facilitates comparative analysis. Moreover, the lake supports economically and ecologically important species, including numerous threatened taxa

(Fujioka et al. 2024), underscoring the need for robust biodiversity monitoring tools. Given these factors, Lake Biwa represents an ideal testbed for evaluating eDNA metabarcoding methodologies using nuclear markers. We developed primers for the ITS2 region, collected reference sequences, and assessed their species-level resolution in comparison to the MiFish region.

Materials and methods

Designing primers

Sequences containing the 5.8S to 28S regions flanking the ITS2 in fish nuclear ribosomal RNA genes were collected from the database of National Center for Biotechnology Information (NCBI) and aligned using MUSCLE (Edgar 2004). Primers were designed to have Tm values of approximately 60 °C for both primers. To ensure that the full length of the ITS2 region was amplified, the forward primer was placed in the 5.8S region and the reverse primer in the 28S region. These regions were chosen because they contain highly conserved sequences across diverse fish taxa, which we confirmed by visually inspecting multiple sequence alignments. Specifically, the forward primer was designed based on sequences from 113 species belonging to 64 genera, 36 families and 12 orders, and the reverse primer was designed based on sequences from 96 species belonging to 58 genera, 34 families, and 11 orders. No primer design software was used; instead, conserved sequences were identified manually to maximize coverage across taxa. Mismatches between each sequence and primer were identified (Table 1).

In vitro checking using tissue derived DNA

To validate the designed primers, polymerase chain reactions (PCRs) were performed using tissue derived DNA of 12 species inhabiting Lake Biwa (*Anguilla japonica*, *Cyprinus carpio*, *Acheilognathus rhombeus*, *Silurus asotus*, *Pelteobagrus nudiceps*, *Lepomis macrochirus*, *Micropterus nigricans*, *Micropterus dolomieu*, *Odontobutis obscura*, *Rhinogobius flumineus*, *Gymnogobius urotaenia*, and *Gymnogobius isaza*). The composition of PCR reagent was 1×Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA), 500 nM each of primers, 1 µl DNA and ultrapure water to a total solution volume of 20 µl. Temperature conditions were 98 °C for 30 s followed by 40 cycles of 98 °C for 10 s, 60 °C for 10 s, 72 °C for 30 s and final extension at 72 °C for 5 min. Electrophoresis was performed using 1.5% gel to check for the presence of bands.

Reference sample collection for metabarcoding

The target species for reference sequences collection were primarily selected based on Fujioka et al. (2024), which provided a comprehensive list of species inhabiting Lake Biwa. Fujioka et al. (2024) distinguishes between the Japanese strain of Common carp (*Cyprinus* sp.) and Common carp (*Cyprinus* spp.). However, it is difficult to compare the two species because there has been insufficient research on whether they can be distinguished using eDNA, particularly with MiFish. Therefore, following the conventional classification of Nakabo (2013), we treated them as a single species, *Cyprinus carpio*.

In addition to species inhabiting Lake Biwa, we included *Mugil cephalus* and *Lateolabrax japonicus*, which are distributed in the Yodo River, downstream of the lake. As a result, a total of 84 fish species were selected for reference sequence collection (Suppl. material 2). Samples were primarily obtained from the Lake Biwa and Yodo River basins. However, for certain species, including non-native species, specimens were also collected from areas outside Lake Biwa (e.g., Gifu and Kanagawa Prefectures) (Suppl. material 2). Samples were collected by the authors using dip nets and cast nets, and were also obtained through donations from the Lake Biwa Museum and freshwater fish researchers, as well as purchases from aquarium shops and fishery cooperative in Shiga Prefecture.

We used specimens maintained at the Lake Biwa Museum for species that cannot be readily collected in the wild. These included protected species, such as *Acheilognathus cyanostigma* and *Gasterosteus aculeatus* subsp., which require governmental permission and endangered species such as *Cobitis minamorii oumiensis*.

DNA extraction and amplification

Total genomic DNA was extracted from a portion of the tissues or pelvic fin of 84 fish species using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Total genomic DNA was used to amplify DNA fragments using PCR. The primer pair developed in this study was used for PCR. The 1st PCR was performed in a 12 µl reaction volume containing 0.24 µl of KOD Fx Neo (Toyobo, Osaka, Japan), 6 µl of 2 × Buffer for KOD Fx Neo, 0.4 mM dNTPs, each of 0.3 µM Fish_ITS2_F/R, and 2 µl of template DNA. The thermal cycle profile was 94 °C for 2 min; 35 cycles of 98 °C for 10 s, 63 °C for 30 s, and 68 °C for 30 s; and 68 °C for 5 min. A PCR non-template control (PCR-NTC) was included using Ultrapure DNase/RNase-free distilled water (ThermoFisher Scientific, Waltham, USA). The PCR amplicons were treated with ExoSAP-IT (ThermoFisher Scientific, Waltham, USA). The PCR amplicons were added to each sample in 1.08 µl of ultrapure water and 0.12 µl of ExoSAP-IT and treated at 37 °C for 15 min and 80 °C for 15 min. In the second PCR, primers (index primer) were used that had MiSeq adaptor sequences and eight bp index sequences (Hamady et al. 2008) to both amplicon ends. The total reaction volume of the 2nd PCR was 13 µl containing 0.24 µl of KOD Fx Neo, 6 µl of 2 × Buffer for KOD Fx Neo, 0.4 mM dNTP, each of 0.5 µM index primer, 1 µl of ultrapure water, and 1 µl of the 1st PCR amplicons. The thermal cycle profile was 95 °C for 3 min; 10 cycles of 98 °C for 20 s, 65 °C for 15 s, and 72 °C for 15 s; and 72 °C for 5 min. A 2% agarose gel electrophoresis was performed to confirm the amplification of the 2nd PCR amplicons.

Amplicon library construction and MiSeq sequencing

The samples were aggregated in a volume of 3 µl per sample based on the electrophoresis results. AMPure XP (Beckman Coulter, Indianapolis IN, USA) was added to 0.6 µl per 1 µl of each sample and treated according to the manual. DNA was eluted with 60 µl of TE buffer. The collected DNA concentration was measured using a Qubit dsDNA HS assay kit (Thermo Fisher Scientific,

Waltham, MA, USA). Finally, a Qsep1 Bio-Fragment Analyzer (BiOptic Inc., Taipei, Taiwan) was used to confirm that the PCR amplicons were of the desired size.

The libraries were sequenced using an Illumina MiSeq with a reagent kit v3 for 2 × 300 bp PE (Illumina, San Diego, CA, USA). Raw reads were demultiplexed into each sample by a preinstalled program in MiSeq, and the outputted fastq files were processed by DADA2 (version 1.20.0; Callahan et al. 2016) as its core program as follows. 1) Primer sequences were removed from both ends of the reads using filterAndTrim function. 2) ASV sequences were inferred using dada function. 3) Forward and reverse reads were merged using mergePairs function. Finally, ASV sequence table was outputted as a csv file and used for the analysis hereafter.

The species assignment for each ASV was conducted through the following steps, and a single ASV was determined for each sample. 1) ASVs obtained were compared using BLAST, and those showing ≥ 80% sequence similarity to species belonging to a different order were likely contaminants and were removed. 2) ASVs with high sequence similarity to species belonging to a different order were also removed during the phylogenetic analysis described later. 3) Considering possible sequencing errors, when multiple ASVs were obtained from a single sample and these ASVs differed by only a few base pairs (approximately 1–3 bp), only the ASV with the highest read count was retained.

Two samples (*Tanakia limbata* and *T. lanceolata*) were sequenced by Sanger sequencing. The DNA extraction and PCR protocols were the same as those for the extraction and 1st PCR described above. As for the primers used in PCR, only the region of interest-specific primers from Fish_ITS2_F/R (see Results) were used. Sequencing was performed for each F/R primer, and the commercial sequencing service (Eurofins Genomics, Inc.) was used.

All sequences were deposited in the DNA Data Bank of Japan (DDBJ), European Nucleotide Archive (EMBL), and GenBank databases under the accession numbers LC876517–LC876631.

Definition of species distinguishability

Among the freshwater fish inhabiting the Lake Biwa basin, some species that are generally considered identifiable using MiFish may, in certain studies, be conservatively assigned only to the genus level due to limited genetic differentiation between species (Nakagawa et al. 2018; Ito et al. 2023; Biodiversity Center of Japan 2024). For example, within *Oncorhynchus* (Salmonidae), *O. biwaensis*, *O. masou ishikawae*, and *O. mykiss* exhibit closely related sequences. Similarly, in *Cobitis* (Cobitidae), *Cobitis* sp. BIWAE type A, *Cobitis* sp. BIWAE type B, *C. magnostriata,* and *C. minamorii oumiensis* also show closely related sequences. In *Carassius* (Cyprinidae), *C. cuvieri* can be distinguished, whereas closely related sequences are observed between other *Carassius* species (*C. buergeri grandoculis* and *Carassius* sp.). Similarly, in *Rhinogobius* (Gobiidae), *R. flumineus* can be distinguished, whereas closely related sequences are observed between other *Rhinogobius* species (*R. biwaensis* and *Rhinogobius* sp. OM). Additionally, MiFish cannot fully distinguish either between *C. buergeri grandoculis* and *Carassius* sp. or between *R. biwaensis* and *Rhinogobius* sp. OM (Biodiversity Center of Japan 2024).

Therefore, we focused on four genera and evaluated species-level distinguishability based on sequence data using BLAST percent identity. BLAST was conducted using the ITS2 region sequences obtained in this study, and percent identity values were extracted among them. For comparison, BLAST was also performed for the MiFish region of the same species, and the corresponding percent identity values were similarly extracted. To compare the obtained percent identity values, a figure showing the maximum and minimum values for interspecific comparisons was created using ggplot2 package in R version 4.3.1 (R Core Team 2023). In this study, referring to the MiFish identity threshold (98.5% at Biodiversity Center of Japan 2024), species were classified into the following four categories:

- Easily distinguishable species: Interspecific identity is ≤ 98.5% and lower than the MiFish identity.
- Distinguishable species: Interspecific identity is ≤ 98.5% and comparable to the MiFish identity.
- Difficult-to-distinguish species: Interspecific identity is ≥ 98.5% or largely overlaps with intraspecific identity.
- Indistinguishable species: Different species share identical sequences, resulting in 100% identity.

For the other fish species, phylogenetic trees were constructed to illustrate interspecific differences. Multiple sequence alignments were performed using MAFFT version 7 (Katoh et al. 2019). Phylogenetic trees were generated for each taxonomic group at the order, family, and genus level. Several sequences were obtained from GenBank (Suppl. material 2). Phylogenetic analyses were conducted using the maximum likelihood (ML) method implemented in MEGA 11 (Tamura et al. 2021), applying the model selected for each tree based on the Bayesian information criterion (BIC). The reliability of each internal branch was assessed using a bootstrap test with 1000 replicates. Partial deletion was applied to handle gaps and missing data.

Results

Designed primers

The primers designed were as follows (Forward Primer: Fish_ITS2_F, 5’-CYACGCCTGTCTGAGGGTC-3’, Reverse Primer: Fish_ITS2_R, 5’-TCTCGTCTGATCTGAGGTCG -3’). In the forward primer, mismatches were identified at 10 of 19 bases; for the 18th base from the 3’ side, the number of fish species with mismatches was high (C: 20, T: 93 species) and was designated as C or T (expressed as Y). Other bases had fewer mismatches, only for 1–6 species among 113 species (Table 1). In the reverse primer, mismatches were identified in 6 of 20 bases. The number of fish species with base mismatches was small, ranging from 2–6 species among 96 species (Table 1). With regard to the length of the amplified regions of the taxa present in the study area, the longest amplified region was 481 bp for *Micropterus nigricans* and the shortest was 298 bp for *Channa argus*. PCR with DNA extracted from the tissues using the designed primers showed bands for every species.

Table 1. Mismatches between the designed primers and collected fish sequences. Grey indicates sequences with mismatches.

Forward Primer

	5'-	C	Y	A	C	G	C	C	T	G	T	C	T	G	A	G	G	G	T	C	-3'	
C		112	20	1	113		107	111			1	112					2			113		
G						112		1		113				112		113	111	113				
A		1		112		1		1			2			1	113							
T			93				6		113		110	1	113						113			

Reverse Primer

	5'-	T	C	T	C	G	T	C	T	G	A	T	C	T	G	A	G	G	T	C	G	-3'
C			93		94		6	92				3	94							96		
G						96				96					96		96	96			96	
A											96					96						
T		96	3	96	2		96	4	96			93	2	96					96			

Reference sequences

In this study, a total of 12,651,181 reads were obtained from the two MiSeq runs. We analyzed 84 fish species inhabiting the Lake Biwa basin and successfully obtained ITS2 sequences for 61 of them, comprising 291,855 reads. Additionally, sequences for two species (*Cyprinus carpio* and *Plecoglossus altivelis altivelis*) could not be obtained in this study because target sequences were not recovered due to contamination with sequences derived from other organisms such as bacteria; however, sequence data for these species were available from previous studies. As a result, we obtained reference ITS2 sequences for a total of 63 fish species in the Lake Biwa basin, accounting for 75% of the target species. Sequences could not be obtained for the remaining 21 species due to sample unavailability (3 species) or poor sequencing quality (18 species) (Suppl. material 2).

Percent identity values between each target species and its closely related species were extracted. Details for the four focal genera are described below, while other taxa are listed in Suppl. material 3.

Phylogenetic trees were constructed using the obtained sequences, which included the full length of the ITS2 region and approximately 15 bp of the adjacent 5.8S and 28S regions (Suppl. material 1: figs S1–S9).

Oncorhynchus

For the three species distributed in the Lake Biwa basin—*Oncorhynchus biwaensis*, *O. masou ishikawae*, and *O. mykiss*—percent identity values with reference sequences were examined using BLAST (Fig. 1). As a result, *O. biwaensis* showed percent identities 1.9–2.8% lower in the ITS2 region than those obtained using MiFish when compared with *O. m. ishikawae* and *O. mykiss*. Based on these results, *O. biwaensis* was considered an "easily distinguishable species" from the other two species. However, when BLAST was conducted using the *O. biwaensis* sequences, the intraspecific identities of *O. biwaensis* largely overlapped with the interspecific identities between *O. biwaensis* and the other two species. Between *O. m. ishikawae* and *O. mykiss*, the interspecific identity values were ≤ 98.5%, but comparable to those obtained using MiFish; therefore, these two species were classified as "distinguishable species" (Suppl. material 3).

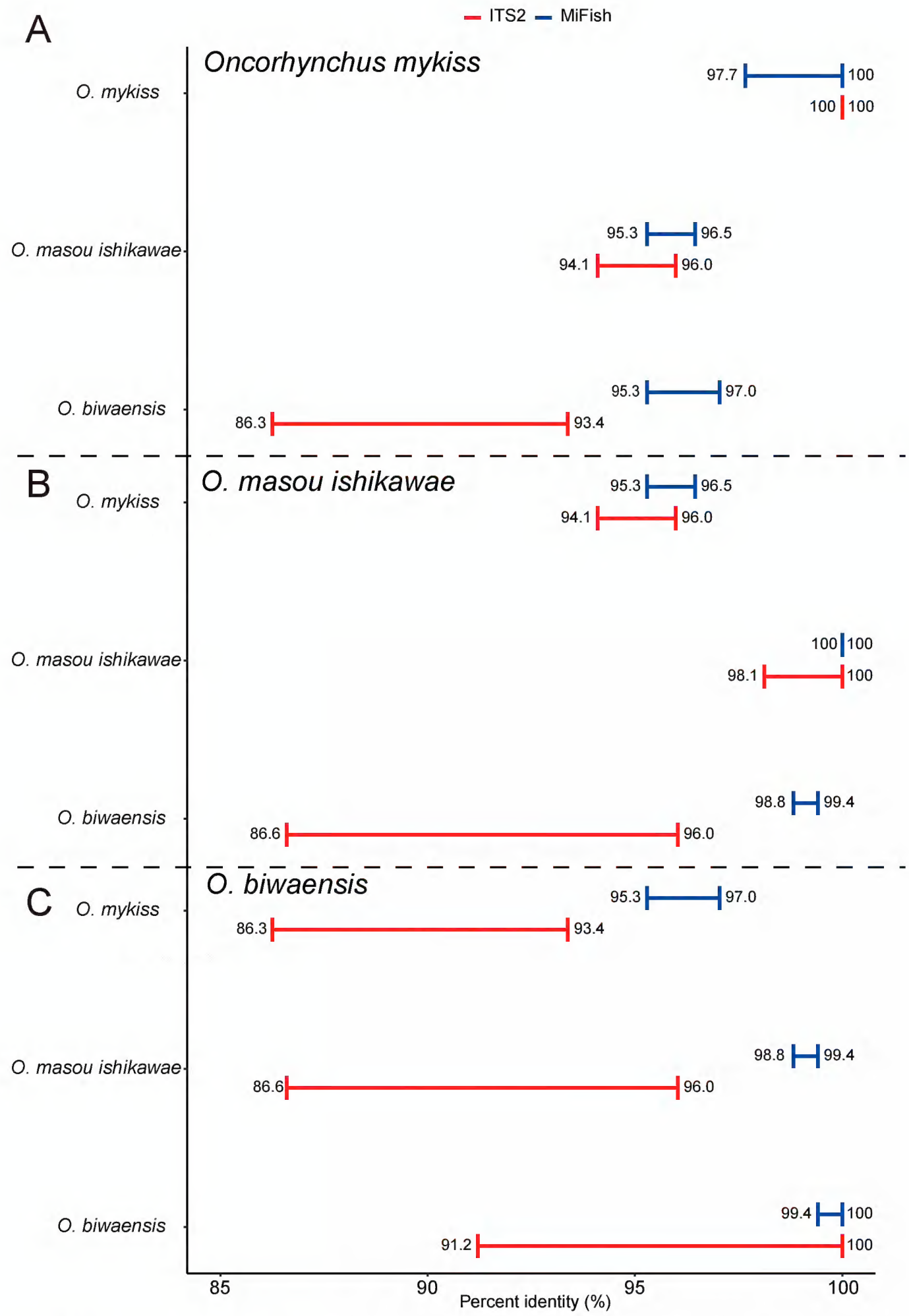

Figure 1. Range between the maximum and minimum percent identity values of ITS2 and MiFish region sequences in *Oncorhynchus*. The x-axis indicates percent identity, and the y-axis shows the reference sequences. Red bars represent the percent identity range for the ITS2 region, and blue bars represent that for the MiFish region. **A** Results when *O. mykiss* was used as the query sequence. **B** Results when *O. masou ishikawae* was used as the query sequence. **C** Results when *O. biwaensis* was used as the query sequence.

Rhinogobius

For the three species inhabiting the Lake Biwa basin—*Rhinogobius* sp. OM, *R. biwaensis*, and *R. flumineus*—percent identity values were examined using BLAST (Fig. 2). As a result, *R. flumineus* showed percent identities 1.0–3.0% lower in the ITS2 region than those obtained using MiFish when compared with

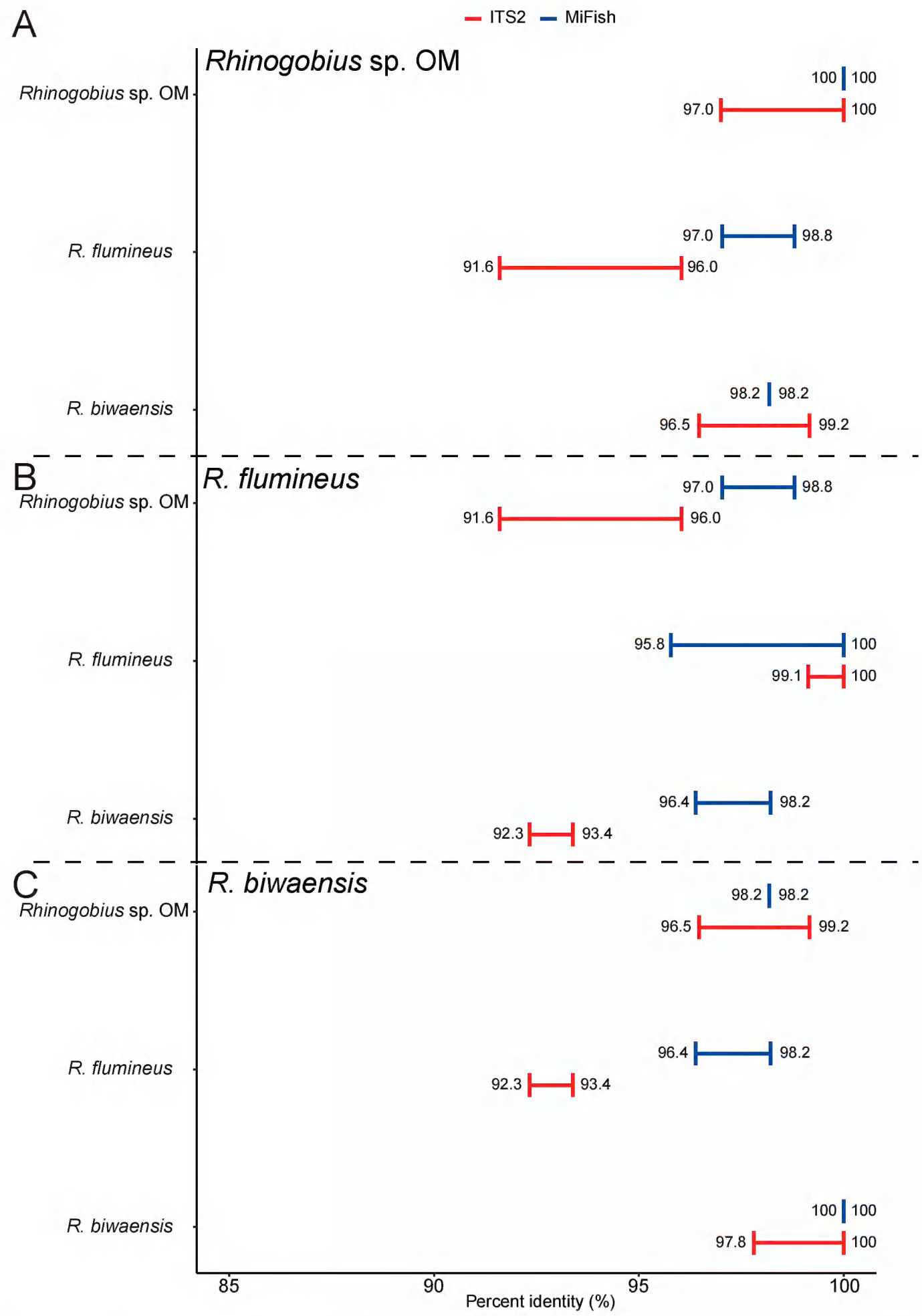

Figure 2. Range between the maximum and minimum percent identity values of ITS2 and MiFish region sequences in *Rhinogobius*. The x-axis indicates percent identity, and the y-axis shows the reference sequences. Red bars represent the percent identity range for the ITS2 region, and blue bars represent that for the MiFish region. **A** Results when *Rhinogobius* sp. OM was used as the query sequence. **B** Results when *R. flumineus* was used as the query sequence. **C** Results when *R. biwaensis* was used as the query sequence.

Rhinogobius sp. OM and *R. biwaensis*. Based on these results, *R. flumineus* was considered an "easily distinguishable species" from the other two species. Between *Rhinogobius* sp. OM and *R. biwaensis*, the percent identity value exceeded 99.5% and largely overlapped with their respective intraspecific identities; therefore, these two species were classified as "difficult-to-distinguish species" (Suppl. material 3).

Cobitis

For the four species inhabiting the Lake Biwa basin—*Cobitis* sp. BIWAE type A, *Cobitis* sp. BIWAE type B, *C. magnostriata*, and *C. minamorii oumiensis*—percent identity values were examined using BLAST (Fig. 3). As a result, *C. magnostriata* and *C. m. oumiensis* showed identical ITS2 region sequences (100% identity), and thus were classified as "indistinguishable species." Between *Cobitis* sp. BIWAE type A and *Cobitis* sp. BIWAE type B, percent identity values exceeded 99.5% and largely overlapped with their respective intraspecific identities; therefore, these two species were judged to be "difficult-to-distinguish species." In contrast, between *Cobitis* sp. BIWAE type A or *Cobitis* sp. BIWAE type B and *C. magnostriata* or *C. m. oumiensis*, the interspecific identity values were ≤ 98.5%, but comparable to those obtained using MiFish; therefore, each pair was classified as "distinguishable species" (Suppl. material 3).

Carassius

For the three species inhabiting the Lake Biwa basin—*Carassius* sp., *C. cuvieri*, and *C. buergeri grandoculis*—percent identity values were examined using BLAST (Fig. 4). As a result, *Carassius* sp. and *C. b. grandoculis* showed identical ITS2 region sequences (100% identity), and thus were classified as "indistinguishable species." In addition, between *C. cuvieri* and *Carassius* sp. or *C. b. grandoculis*, the percent identity values exceeded 98.5%; therefore, species were classified as "difficult-to-distinguish species" (Suppl. material 3).

Discussion

Utility of ITS2 in species discrimination

This study developed primers for the ITS2 region and provided reference sequences for 63 fish species in the Lake Biwa basin, including sequences registered in public databases. One key advantage of ITS2 eDNA metabarcoding is its potential to resolve species that are difficult to distinguish using MiFish. A notable example is *Oncorhynchus biwaensis*, which was judged to be more easily distinguishable from *O. masou ishikawae* and *O. mykiss* based on percent identity than when using the MiFish. While *O. biwaensis* and *O. m. ishikawae* have been shown to be distinct based on nDNA AFLP analysis—a method that examines multiple loci (Kuwahara et al. 2012)—this study suggests that species discrimination may be feasible even with a single ITS region. Given that *O. biwaensis* is an important fishery resource (Fujioka et al. 2024), applying this marker to eDNA surveys could provide valuable insights into its precise distribution and contribute to fisheries management and ecological studies. However, considering that *O. biwaensis* in Lake Biwa hybridizes with the introduced *O. m. ishikawae* (Kuwahara et al. 2012), if DNA from both species is detected in sympatric habitats, it may indicate the detection of DNA derived from hybrid individuals. Nevertheless, it should be noted that detecting hybrids using eDNA ideally requires single cell based eDNA analysis, and the current methods cannot rigorously determine whether the detected DNA originates from hybrids.

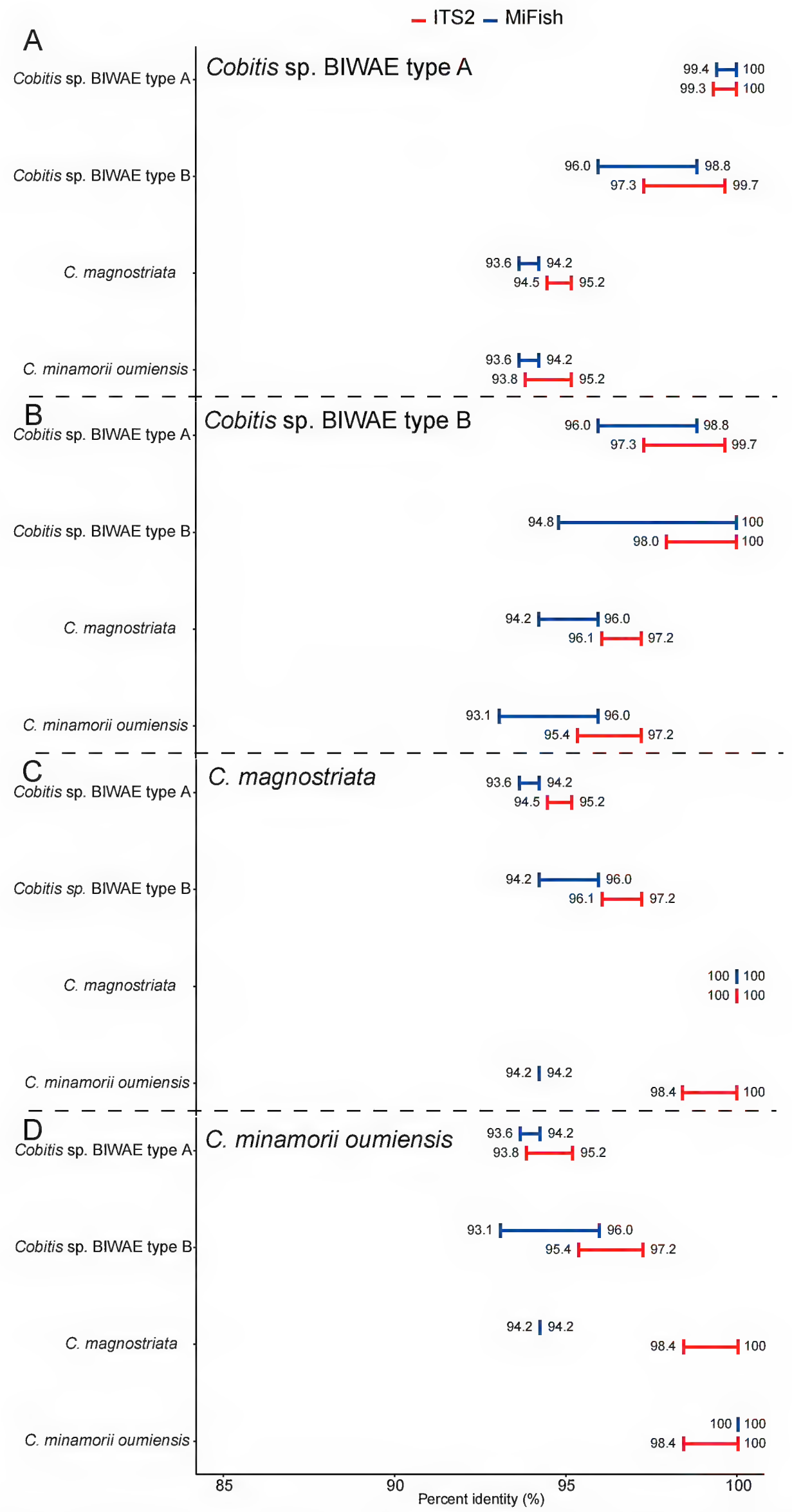

Figure 3. Range between the maximum and minimum percent identity values of ITS2 and MiFish region sequences in *Cobitis*. The x-axis indicates percent identity, and the y-axis shows the reference sequences. Red bars represent the percent identity range for the ITS2 region, and blue bars represent that for the MiFish region. **A** Results when *Cobitis* sp. BIWAE type A was used as the query sequence. **B** Results when *Cobitis* sp. BIWAE type B was used as the query sequence. **C** Results when *C. magnostriata* was used as the query sequence. **D** Results when *C. minamorii oumiensis* was used as the query sequence.

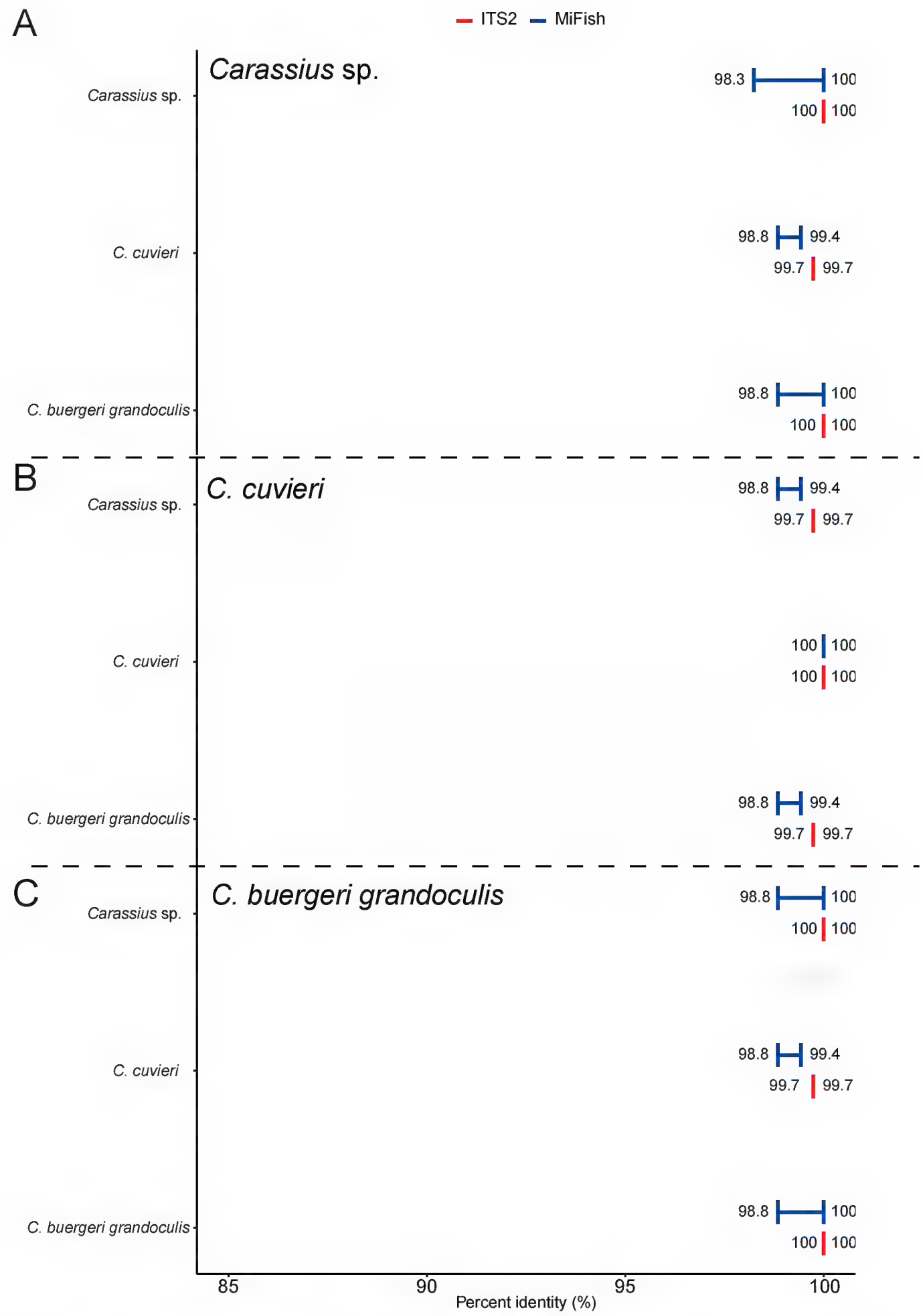

Figure 4. Range between the maximum and minimum percent identity values of ITS2 and MiFish region sequences in *Carassius*. The x-axis indicates percent identity, and the y-axis shows the reference sequences. Red bars represent the percent identity range for the ITS2 region, and blue bars represent that for the MiFish region. **A** Results when *Carassius* sp. was used as the query sequence. **B** Results when *C. cuvieri* was used as the query sequence. **C** Results when *C. buergeri grandoculis* was used as the query sequence.

In the case of *O. biwaensis*, the intraspecific percent identity values largely overlapped with the interspecific values between *O. masou ishikawae* and *O. mykiss*. This suggests that one ITS2 sequence of *O. biwaensis* (LC876601) may exhibit greater genetic variation within the species than that observed between species. However, it is also possible that LC876601 represents an erroneous sequence generated by sequencing errors. Because only two individuals of *O. biwaensis* were analyzed in this study, it is difficult to determine whether

this sequence is authentic. Further analyses using a larger number of *O. biwaensis* individuals are needed to clarify the extent of intraspecific polymorphism in the ITS2 region.

Rhinogobius flumineus was also judged to be more easily distinguishable from *Rhinogobius* sp. OM and *R. biwaensis* based on percent identity than when using the MiFish. Molecular phylogenetic analyses using nDNA have shown that this species diverged earlier (approximately 4.48 MYA) than other *Rhinogobius* species, including *Rhinogobius* sp. OM and *R. biwaensis* (Yamasaki et al. 2015). The deep divergence time in nDNA likely explains why *R. flumineus* was more easily distinguishable than the other two species. This species tends to inhabit upstream riverine areas within the Lake Biwa basin (Fujioka et al. 2024) and may face population declines due to habitat modification in these upper reaches. Therefore, the ability to reliably distinguish this species holds importance for its conservation.

Reasons for indistinguishable or difficult-to-distinguish species

This study identified four fish species classified as "indistinguishable species" based on the ITS2 region. In *Cobitis*, *C. magnostriata*, which was indistinguishable from *C. minamorii oumiensis*, is known to be an allotetraploid species derived from at least three parental lineages, including *C. minamorii* (Saitoh et al. 2010). Due to this polyploid origin, *C. magnostriata* and *C. m. oumiensis* share identical sequences across several nDNA regions (Saitoh et al. 2010), suggesting that they could not be distinguished using the nDNA-based ITS2 region.

In *Carassius*, *Carassius* sp. and *C. buergeri grandoculis* were indistinguishable based on the ITS2 region. These two species are known to be difficult to distinguish using both mtDNA and nDNA markers (Yamamoto et al. 2010; Mishina et al. 2021). *Carassius* sp. is a gynogenetic triploid lineage that occasionally hybridizes with diploid species of *Carassius*, including *C. b. grandoculis* (Yamamoto et al. 2010; Mishina et al. 2021). Such genetic homogeneity and ongoing gene flow are likely responsible for the inability to distinguish the two species using the ITS2 region.

Three species were also classified as "difficult-to-distinguish species" based on the ITS2 region. In *Rhinogobius*, *Rhinogobius* sp. OM and *R. biwaensis*, both distributed within the Lake Biwa basin, were considered difficult to distinguish. These two species, together with *R. tyoni*, are closely related based on nDNA and exhibit a relatively shallow divergence time (approximately 1.10 MYA). Furthermore, complex hybridization events among these three species have been suggested (Fujioka et al. 2024). This genetic homogeneity and ongoing gene flow likely resulted in nearly identical ITS2 sequences between *Rhinogobius* sp. OM and *R. biwaensis*.

The closely related species *Cobitis* sp. BIWAE type A and *Cobitis* sp. BIWAE type B were also classified as difficult to distinguish. The former is considered an autopolyploid derivative of the latter (Kimizuka and Kobayashi 1983; Saitoh et al. 2010), which accounts for their close phylogenetic relationship. However, their distributions are almost entirely allopatric (Kimizuka and Kobayashi 1983), indicating that incorporating geographic information can aid species identification in field surveys.

Carassius cuvieri was judged to be difficult to distinguish from *Carassius* sp. and *C. b. grandoculis*. Although *C. cuvieri* is generally considered clearly distinguishable from other *Carassius* species, including *Carassius* sp. and *C. b. grandoculis*, based on both mtDNA and nDNA markers (Yamamoto et al.

2010; Mishina et al. 2021), these three species exhibited nearly identical ITS2 sequences. This unexpected result may reflect a slower evolutionary rate of the ITS2 region in *C. cuvieri*, and the use of additional nDNA loci may improve the accuracy of species identification in future studies.

Species requiring additional reference data

This study was unable to obtain sequences for species from the families Anguillidae, Botiidae, Adrianichthyidae, Poeciliidae, Sinipercidae, and Synbranchidae. Since only one species from each of these families inhabits Lake Biwa, and no closely related species have been sequenced, identification via eDNA metabarcoding remains challenging. Thus, obtaining reference sequences for these species should be prioritized in future research.

For the remaining 14 fish species, sequences from closely related taxa were available, potentially allowing species inference based on phylogenetic relationships. However, careful evaluation is required to determine whether these species can be reliably distinguished. Key taxa of interest include *Acheilognathus* (*A. tabira tabira* and *A. longipinnis*), *Biwia yodoensis*, *Hemibarbus labeo*, *Squalidus japonicus japonicus*, *Silurus lithophilus*, and *Gymnogobius isaza*. Among these, *Hemibarbus*, *Squalidus*, and *Silurus* include species that are indistinguishable using MiFish. If ITS2 can differentiate them, it would provide further justification for incorporating ITS2-based eDNA metabarcoding into biodiversity assessments.

Conclusion

This study compiled reference ITS2 sequences for freshwater fish species in the Lake Biwa basin and demonstrated their utility in species identification. The findings suggest that ITS2 can serve as an effective eDNA metabarcoding marker, particularly for species that are difficult to distinguish using MiFish. Notably, *O. biwaensis* exhibited greater phylogenetic resolution with ITS2 than with MiFish, highlighting its potential for species discrimination. On the other hand, for *C. magnostriata* and *C. m. oumiensis*, which were classified as indistinguishable, as well as for *C. cuvieri*, *Carassius* sp., and *C. b. grandoculis*, which were classified as difficult-to-distinguish, MiFish demonstrated superior discriminative ability (Biodiversity Center of Japan 2024). This finding also indicates that a single genetic marker does not necessarily yield optimal results for all taxa.

To fully capture biodiversity information, a combined approach utilizing both ITS2 and MiFish is recommended. Future research should focus on validating ITS2-based eDNA metabarcoding in field applications within the Lake Biwa basin and expanding the reference database to improve taxonomic resolution. By incorporating additional species and individuals, this approach can enhance the accuracy and applicability of eDNA metabarcoding for fish biodiversity monitoring.

Acknowledgments

We express sincere thanks to Dr. Maki Asami, Ms. Yuko Goto, and Mr. Rintaro Kimura for their help with DNA analysis.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

The following AI tools were used in the preparation of this manuscript: Description: For language editing, ChatGPT based on the GPT-4-turbo (OpenAI 2024) was employed to identify and correct grammatical inaccuracies and imprecise phrasing.

Funding

This study was supported by JSPS KAKENHI (JP20H03326 and JP23H02556).

Author contributions

Gen Ito: Investigation, Data curation, Formal Analysis, and Writing – original draft. Daisuke Sasaki: Formal Analysis and Writing – review & editing, Masayuki K. Sakata: Data curation and Writing – review & editing, Seigo Kawase: Resources and Writing – review & editing, Toshifumi Minamoto: Conceptualization, Data curation, Funding acquisition, and Writing – review & editing, Hiroki Yamanaka: Supervision, Conceptualization, Data curation, Funding acquisition, and Writing – review & editing.

Author ORCIDs

Gen Ito https://orcid.org/0000-0002-9781-7206
Masayuki K. Sakata https://orcid.org/0000-0003-4837-854X
Toshifumi Minamoto https://orcid.org/0000-0002-5379-1622
Hiroki Yamanaka https://orcid.org/0000-0002-6193-1656

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular Biology of the Cell, 4th edn. Garland Science, New York, 712 pp.

Avise JC, Lansman RA, Shade RO (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. Genetics 92: 279–295. https://doi.org/10.1093/genetics/92.1.279

Barbarossa V, Bosmans J, Wanders N, King H, Bierkens MF, Huijbregts MA, Schipper AM (2021) Threats of global warming to the world's freshwater fishes. Nature Communications 12: 1701. https://doi.org/10.1038/s41467-021-21655-w

Biodiversity Center of Japan (2024) Misidentification checksheet for MiFish metabarcoding ver.1.2. [April 7, 2025] https://www.biodic.go.jp/edna/edna_top.html

Booton GC, Kaufman L, Chandler M, Oguto-Ohwayo R, Duan W, Fuerst PA (1999) Evolution of the ribosomal RNA internal transcribed spacer one (ITS-1) in cichlid fishes of the Lake Victoria Region. Molecular Phylogenetics and Evolution 11: 273–282. https://doi.org/10.1006/mpev.1998.0576

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–97. https://doi.org/10.1093/nar/gkh340

Fujioka Y, Kawase S, Tabata R (2024) The Fishes of Lake Biwa. Sunrise Publishing, Hikone, 229 pp.

Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25: 3101–3119. https://doi.org/10.1111/mec.13660

Hamady M, Walker JJ, Harris K, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nature Methods 5: 235–237. https://doi.org/10.1038/nmeth.1184

Hidaka S, Jo TS, Yamamoto S, Katsuhara KR, Tomita S, Miya M, Ikegami M, Ushimaru A, Minamoto T (2024) Sensitive and efficient surveillance of Japanese giant salamander (*Andrias japonicus*) distribution in western Japan using multi-copy nuclear DNA marker. Limnology 25: 189–198. https://doi.org/10.1007/s10201-023-00740-7

Huang LL, Wu ZQ, Li JH (2013) Fish fauna, biogeography and conservation of freshwater fish in Poyang Lake Basin, China. Environmental Biology of Fishes 96: 1229–1243. https://doi.org/10.1007/s10641-011-9806-2

Ito G, Yamauchi H, Shigeyoshi M, Ashino K, Yonashiro C, Asami M, Goto Y, Duda JJ, Yamanaka H (2023) Using eDNA metabarcoding to establish targets for freshwater fish composition following river restoration. Global Ecology and Conservation 43: e02448. https://doi.org/10.1016/j.gecco.2023.e02448

Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M, Nishida M (2013) MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accu- rate and automatic annotation pipeline. Molecular Biology and Evolution 30: 2531–2540. https://doi.org/10.1093/molbev/mst141

Kano Y, Musikasinthorn P, Iwata A, Tun S, Yun LKC, Win SS, Matsui S, Tabata R, Yamasaki T, Watanabe K (2016) A dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models. Biodiversity Data Journal 4: e10539. https://doi.org/10.3897/BDJ.4.e10539

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20: 1160–1166. https://doi.org/10.1093/bib/bbx108

Kimizuka Y, Kobayasi H (1983) Geographic distributions of karyological races of *Cobitis biwae* (Cobitididae). Japanese Journal of Ichthyology 30: 308–312. https://doi.org/10.11369/jji1950.30.308

Kuwahara M, Takahashi H, Kikko T, Kurumi S, Iguchi K (2012) Introgression of *Oncorhynchus masou* subsp. (Biwa salmon) genome into lake-run *O. m. ishikawae* (Amago salmon) introduced into Lake Biwa, Japan. Ichthyological Research 59: 195–201. https://doi.org/10.1007/s10228-011-0270-y

Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annual Review of Biochemistry 49: 727–764. https://doi.org/10.1146/annurev.bi.49.070180.003455

Minamoto T, Uchii K, Takahara T, Kitayoshi T, Tsuji S, Yamanaka H, Doi H (2017) Nuclear internal transcribed spacer-1 as a sensitive genetic marker for environmental DNA studies in common carp *Cyprinus carpio*. Molecular Ecology Resources 17: 324–333. https://doi.org/10.1111/1755-0998.12586

Mishina T, Takeshima H, Takada M, Iguchi K, Zhang C, Zhao Y, Kawahara-Miki R, Hashiguchi Y, Tabata R, Sasaki T, Nishida M, Watanabe K (2021) Interploidy gene flow involving the sexual-asexual cycle facilitates the diversification of gynogenetic triploid *Carassius* fish. Scientific Reports 11: 22485. https://doi.org/10.1038/s41598-021-01754-w

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science 2: 150088. https://doi.org/10.1098/rsos.150088

Nakabo T (2013) Fishes of Japan with pictorial keys to the species, 3rd edn. Tokai University Press, Hadano, 2428 pp.

Nakagawa H, Yamamoto S, Sato Y, Sado T, Minamoto T, Miya M (2018) Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods. Freshwater Biology 63: 569–580. https://doi.org/10.1111/fwb.13094

R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://doi.org/10.32614/r.manuals

Saitoh K, Chen WJ, Mayden RL (2010) Extensive hybridization and tetrapolyploidy in spined loach fish. Molecular Phylogenetics and Evolution 56: 1001–1010. https://doi.org/10.1016/j.ympev.2010.04.021

Sigsgaard EE, Carl H, Møller PR, Thomsen PF (2015) Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. Biological Conservation 183: 46–52. https://doi.org/10.1016/j.biocon.2014.11.023

Su G, Logez M, Xu J, Tao S, Villéger S, Brosse S (2021) Human impacts on global freshwater fish biodiversity. Science 371: 835–838. https://doi.org/10.1126/science.abd3369

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Molecular Biology and Evolution 38: 3022–3027. https://doi.org/10.1093/molbev/msab120

Ushio M, Fukuda H, Inoue T, Kobayashi M, Kishida O, Sato K, Murata K, Nikaido M, Sado T, Sato Y, Takeshita M, Iwasaki W, Yamanaka H, Kondoh M, Miya M (2017) Environmental DNA enables detection of terrestrial mammals from forest pond water. Molecular Ecology Resources 17: e63–e75. https://doi.org/10.1111/1755-0998.12690

Ushio M, Murata K, Sado T, Nishiumi I, Takeshita M, Iwasaki W, Miya M (2018) Demonstration of the potential of environmental DNA as a tool for the detection of avian species. Scientific Reports 8: 4493. https://doi.org/10.1038/s41598-018-22817-5

Wu L, Wu Q, Inagawa T, Okitsu J, Sakamoto S, Minamoto T (2022) Estimating the spawning activity of fish species using nuclear and mitochondrial environmental DNA concentrations and their ratios. Freshwater Biology 68: 103–114. https://doi.org/10.1111/fwb.14012

Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, Minamoto T, Miya M (2017) Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. Scientific Reports 7: 40368. https://doi.org/10.1038/srep40368

Yamamoto G, Takada M, Iguchi K, Nishida M (2010) Genetic constitution and phylogenetic relationships of Japanese crucian carps (*Carassius*). Ichthyological Research 57: 215–222. https://doi.org/10.1007/s10228-010-0152-8

Yamasaki YY, Nishida M, Suzuki T, Mukai T, Watanabe K (2015) Phylogeny, hybridization, and life history evolution of *Rhinogobius* gobies in Japan, inferred from multiple nuclear gene sequences. Molecular Phylogenetics and Evolution 90: 20–33. https://doi.org/10.1016/j.ympev.2015.04.012

Supplementary material 1

Additional figures

Authors: Gen Ito, Daisuke Sasaki, Masayuki K. Sakata, Seigo Kawase, Toshifumi Minamoto, Hiroki Yamanaka

Data type: pdf

Explanation note: **fig. S1.** Maximum likelihood trees of 297 bp ITS2 include partial 5.8S and 28S regions sequences of Petromyzontidae using the T92 model. **fig. S2**. Maximum likelihood trees of 265 bp ITS2 region sequences of Cyprinidae using the GTR+G model. **fig. S3.** Maximum likelihood trees of 290 bp ITS2 include partial 5.8S and 28S regions sequences of Cobitidae using the T92+G model. **fig. S4.** Maximum likelihood trees of 304 bp ITS2 include partial 5.8S and 28S regions sequences of Nemacheilidae using the T92 model. **fig. S5.** Maximum likelihood trees of 202 bp ITS2 include partial 5.8S and 28S regions sequences of Siluridae, Bagridae, Amblycipitidae, and Ictaluridae using the T92+I model. **fig. S6.** Maximum likelihood trees of 299 bp ITS2 include partial 5.8S and 28S regions sequences of Salmonidae using the T92 model. **fig. S7.** Maximum likelihood trees of 281 bp ITS2 include partial 5.8S and 28S regions sequences of Plecoglossidae and Osmeridae using the T92 model. **fig. S8.** Maximum likelihood trees of 221 bp ITS2 include partial 5.8S and 28S regions sequences of Gasterosteidae, Mugilidae, Lateolabracidae, Centrarchidae, Cottidae, and Channidae using the T92+G+I model. **fig. S9.** Maximum likelihood trees of 292 bp ITS2 include partial 5.8S and 28S regions sequences of Odontobutidae and Gobiidae using the T92+G model.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.10.162895.suppl1

Supplementary material 2

Locality information for each sequenced sample

Authors: Gen Ito, Daisuke Sasaki, Masayuki K. Sakata, Seigo Kawase, Toshifumi Minamoto, Hiroki Yamanaka

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.10.162895.suppl2

Supplementary material 3

Distinguishability of ITS2 and MiFish regions in each fish species

Authors: Gen Ito, Daisuke Sasaki, Masayuki K. Sakata, Seigo Kawase, Toshifumi Minamoto, Hiroki Yamanaka

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.10.162895.suppl3